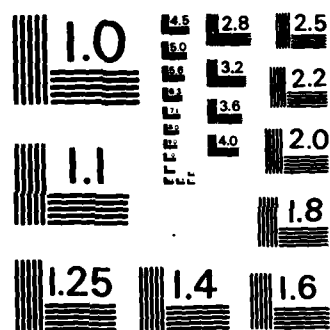


AD-A139 534 A NOTE ON MIXED EXPONENTIAL APPROXIMATIONS FOR GI/G/1 1/1
WAITING TIMES(U) VIRGINIA UNIV CHARLOTTESVILLE DEPT OF
SYSTEMS ENGINEERING C M HARRIS FEB 84
UNCLASSIFIED UVA/525393/SE84/101 N00014-83-K-0624 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

AD A139534

Technical Report
Contract No. N00014-83-K-0624

A NOTE ON MIXED EXPONENTIAL APPROXIMATIONS
FOR GI/G/1 WAITING TIMES

Submitted to:

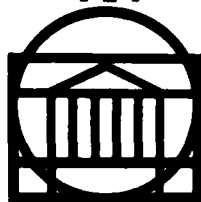
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Mathematical and Physical Sciences

Submitted by:

Carl M. Harris
Professor

Report No. UVA/525393/SE84/101
February 1984



SCHOOL OF ENGINEERING AND
APPLIED SCIENCE

DEPARTMENT OF SYSTEMS ENGINEERING

This document has been approved
for public release and sale; its
distribution is unlimited.

UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22901
84 03 23 032

DTIC
ELECTE
MAR 26 1984
S
E

DTIC FILE COPY

Technical Report

Contract No. N00014-83-K-0624

A NOTE ON MIXED EXPONENTIAL APPROXIMATIONS
FOR GI/G/1 WAITING TIMES

Submitted to:

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Mathematical and Physical Sciences

Submitted by:

Carl M. Harris
Professor

Serial NH-1

Contract N00014-83-K-0624
Task A; Project NR 347-139

Department of Systems Engineering
RESEARCH LABORATORIES FOR THE ENGINEERING SCIENCES
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA

Report No. UVA/525393/SE84/101

February 1984

Copy No. _____

This document has been approved for public sale and release; its distribution is unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AD-A139534		
4. TITLE (and Subtitle) A Note on Mixed Exponential Approximations for GI/G/1 Waiting Times		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER UVA/525393/SE84/101
7. AUTHOR(s) Carl M. Harris		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0624
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Systems Engineering University of Virginia, Thornton Hall Charlottesville, VA 22901		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Task A; Project NR347-139
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE February 1984
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE Unclassified
16. DISTRIBUTION STATEMENT (of this Report) Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) stochastic processes service time approximation statistical inference waiting time computations approximations mixed exponential densities queueing models		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A method is offered for the effective estimation of the stationary waiting-time distribution of the GI/G/1 queue by a (possibly nonconvex) mixed exponential CDF. The approach relies on obtaining a generalized exponential mixture as an approximation for the distribution of the service times. This is done by the adaptation of a nonlinear optimization algorithm previously developed for the maximum-likelihood estimation of parameters from mixed Weibull distributions. The approach is particularly well-suited for obtaining the delay distribution beginning from raw interarrival and service-time data.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

111

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SCOPE AND PURPOSE

It is typically very difficult to obtain waiting-time distribution functions for non-Markovian, general single-server queues. It has long been recognized that virtually any probability density (such as for service times) can be approximated quite accurately by a linear combination of exponential density functions. Fortunately, when such forms can be assumed to describe interarrival and service times, classical results from the theory of queues permit the solution of the waiting-time problem for the general queue. However, there have been no good computer-based algorithms available until recently for determining appropriate mixed exponential densities to fit data or closely approximate another non-standard density. The subject of this work then is the adaptation of a special nonlinear optimization routine for generalized mixture estimation and its use in calculating very accurate approximations for the delay distribution of any general single-server queue.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



I. INTRODUCTION

In a recent paper, Fredericks (1982) offered a class of approximations for the GI/G/1 stationary waiting-time distribution function (CDF), $W_q(t)$. The main idea there was to assume a specific form for $W_q(t)$ and then to use the Lindley integral equation to obtain estimates of the parameters.

The primary form used for the approximant was the exponential

$$W_q(t) = 1 - Ce^{-at}$$

and then numerous ways were documented for the estimation of C and a . However, presetting the form of $W_q(t)$ leads to complications when more complex functional forms are tried, and this approach does not generally portray a waiting-time CDF in an accurate way over its full range. Fredericks also mentioned the mixed exponential as another possible form for $W_q(t)$ in light of its common appearance as a waiting-time distribution.¹ This gave a more precise approximation, but other problems then surfaced.

However, there is an entirely different way to look at the problem which has some precedence in the literature, and which often provides both a special insight into the underlying queue mechanisms and a most accurate way of assessing the waiting times. We suggest looking instead at the interarrival and service-time distributions defining the GI/G/1 as the candidates for the approximation, such that the final form of $W_q(t)$ is a member of a very

¹By mixed exponential, we mean that the complementary CDF $1-W_q(t)$ may be written as a linear combination of exponential functions with possibly complex powers. In its most general form, the (possibly complex) linear combination need not be convex and the constants could be mixed positive and negative. And, of course, the exponents should have negative real parts. See Dehon and Latouche (1982) for an expanded and up-to-date discussion of this class of distribution functions.

comprehensive family of CDFs, namely, the mixed exponentials. So we present a way to approximate which guarantees that the waits turn out to be mixed exponentially distributed.

The primary motivation for our alternative approach is previous work of Smith (1953), and Marchal and Harris (1976). The major relevant result of Smith was his Theorem 4, in which some quite general sufficient conditions were presented for the GI/G/1 waiting times to have a mixed exponential distribution. Marchal and Harris built on this result and some others of Smith to offer a relatively simple mixed-exponential approximation of $W_q(t)$ derived by fitting the difference of the service and interarrival time random variables by a difference of two Erlang variables.

II. NOTATION AND BACKGROUND MATERIAL

The fundamental results for the GI/G/1 may be found, for example, in Gross and Harris (1974), Chapter 6. It was noted there that the GI/G/1 problem can be greatly simplified if it may be assumed that both interarrival and service distributions are generalized Erlangs (GE) expressed as convolutions of independent and not-necessarily-identical exponential random variables. When the means of such exponentials are allowed to come in conjugate pairs (so that their Laplace-Stieltjes transforms are inverse polynomials), Smith calls the family K_n (with n the degree of the defining polynomial). Other authors (e.g., Cohen, 1982) define K_n as the class of distributions whose transforms are rational functions (clearly including the inverse polynomials); but we shall instead call these R_n (with n the degree of the denominator's polynomial). The K_n class includes all regular Erlangs, but not all mixed exponentials and mixed Erlangs, which are however members of R_n . We also mention the generalized phase-type distributions (PH) popularized by Neuts and others (see Neuts, 1981), which have rational transforms as well, though not necessarily of the inverse polynomial form. Thus we may symbolically represent the relationship of those respective families as $GE \subset K_n \subset R_n$ and $PH \subset R_n$.

Now, under a double K_n assumption (i.e., that the queue is $K_m/K_n/1$), it turns out that the delay CDF is (for example, see Gross and Harris)

$$W_q(t) = 1 + \sum_{i=1}^n k_i e^{z_i t}, \quad (1)$$

where the $\{k_i\}$ and $\{z_i\}$ would be determined in the usual way (as in Gross and Harris). The $\{k_i\}$ are arbitrary in sign, while the $\{z_i\}$ have negative real parts. A completely parallel result exists for distributions in R_n .

Importantly, Smith also showed that a comparable result follows even for arbitrary interarrival times. That is, the $GI/K_n/1$ and $GI/R_n/1$ queues have mixed exponential waits independent of the form of GI . And therein lies the key to our approximation method: the interarrival and service CDFs are the functions of concern.

So our approach is built around the approximation of the service distribution by a member of either the class K_n or R_n . Since R_n is the more general class and includes K_n , we focus on selecting from amongst its members. More precisely, we work with a subset of the class R_n made up of those rational functions whose denominators have real roots, and call this class GH for "generalized hyperexponential."²

It is most important to recognize that GH is a very complete set of potential approximants even without the use of complex scale parameters. The most critical characterization of this coverage is the fact that all functions in $L_2(0, \infty)$ can be approximated arbitrarily closely by a finite linear combination of functions of the form $e^{-\phi t}$, $\phi \in \mathbb{R}^+$.³ (For example, see Naylor and Sell, 1971, for a discussion of this and related problems on the Hilbert space of square integrable functions.)

So the distribution selection problem is that of describing the probability structure of the service times by a density of the form

$$b(t) = \sum_{i=1}^K \gamma_i \phi_i e^{-\phi_i t}$$

²Note that the class $GH \supset GE$ and that $GH \subset PH$.

³Indeed, the classes, GH, PH and R_n are each dense in the set of all CDFs on the nonnegative reals.

either from data or as an approximation to an actual available density function whose form is awkward. If we indeed have raw data, then our problem is just one of parameter estimation. Otherwise, we have a curve fitting problem, which in turn can be converted to an estimation problem by picking a large number of the function's points and then proceeding with the maximum-likelihood estimation (MLE). The recommended approach for this selection is based on a nonlinear optimization routine previously developed for the MLE of mixed Weibull parameters. It is presented in the next section.

III. METHOD FOR SERVICE DENSITY APPROXIMATION

The numerical procedure for estimation has been built up from previous work on exponential and Weibull mixtures. To illustrate, let us assume that the data sampling is complete so that all service times are fully observed. In the event that there are incomplete data observations, the algorithm is easily altered.

Maximum-likelihood estimation is the method selected mainly because, under fairly general conditions, it enjoys the important limiting statistical properties of efficiency, normality, and unbiasedness. Furthermore, the MLEs are consistent, invariant, and are functions of sufficient statistics if they exist. When sufficiency and unbiasedness both hold, the MLEs are also of minimum variance.

A first key observation is that it is not possible to obtain explicit formulas for the maximum-likelihood estimators of parameters involved in mixed exponential densities by taking the partial derivatives and equating them to zero. Hence we must resort to other optimization methods and numerical techniques. Furthermore, we need to take into account a set of constraints in addition to the objective function. The mixing proportions and scale parameters must satisfy some simple linear relationships and there may exist other constraints related to the sub-population parameters. Note that the constraints are generally of a linear type; hence the problem can be described as a mathematical program with a nonlinear objective function and linear constraints.

The target criterion function in the maximum-likelihood problem is the joint density function for a random sample from the (mixed) population governed by $b(t)$. As is common, it is much easier in this situation to work

with the logarithm of the likelihood function. Thus if we write the likelihood for the random sample t_1, \dots, t_N as

$$L(\alpha) = \prod_{i=1}^N b(t_i; \alpha) = \prod_{i=1}^N \sum_{j=1}^K \gamma_j \phi_j e^{-\phi_j t_i} \quad (5)$$

where α is the vector of parameters (which may include K), then its logarithm is expressed as

$$\ell(\alpha) = \sum_{i=1}^N \ln b(t_i; \alpha). \quad (6)$$

The general MLE problem for the mixture may then be formulated as the non-linear constrained optimization problem:

$$\max_{\alpha} \ell(\alpha)$$

subject to

$$\alpha \in S = \{\alpha \mid \sum \gamma_j = 1; \phi_j \geq 0\}. \quad (7)$$

Under the standard mixed-exponential regime, all $\gamma_j > 0$, and ϕ_j would be real and greater than 0. The most efficient algorithm available for the solution of this problem is due to the joint efforts of Kaylan (1978) and Kaylan and Harris (1981), and Mandelbaum (1982) and Mandelbaum and Harris (1982). It is an iterative numerical procedure built around Newton's method with additional use of second partial derivatives to speed up convergence whenever necessary.

Because the mixing parameters may be negative in our queueing problem, the Kaylan/Mandelbaum/Harris algorithm has been carefully altered. The basic approach is similar, except that now the number of mixing variables has been doubled. Since it is preferable to optimize over nonnegative variables, we thus write γ_i as the difference between its positive and negative parts. There were two major changes necessary in the algorithm. First, additional

code had to be added to make sure that the density function $b(t)$ did not become negative. A second alteration had to be made to prevent the mixing parameters $\{\gamma_i\}$ from drifting too far out along the real line in either direction.

IV. ILLUSTRATIVE EXAMPLE

To show how our proposed method would work, the following test problem is offered. It is patterned somewhat after an example in Gross and Harris (pp. 300-301) and is of the M/G/1 type. Interarrivals are assumed to have mean 2, while a set of 50 service times has been generated randomly according to an Erlang(2) with mean 1. These numbers are (to two decimal places):

.90,	2.14,	1.33,	.52,	.72,	.06,	.49,	.53,	1.10,	.63
.31,	1.47,	.91,	.26,	.49,	.97,	.24,	2.99,	2.60,	1.11
1.99,	.41,	.31,	1.58,	1.42,	.84,	.11,	.54,	1.27,	2.05
.51,	.91,	1.78,	1.69,	.78,	.64,	1.62,	1.96,	.93,	.41
1.70,	.73,	.98,	.71,	.73,	.44,	1.27,	.92,	.71,	.78

The algorithm of the previous section was then applied to the data and the resultant generalized mixed exponential turned out to be (with its parameters rounded off for simplicity)

$$B(t) = 1 - e^{-4t} + 4e^{-3t} - 4e^{-2t}.$$

The next step is to substitute $B(t)$ into the GI/G/1 waiting-time equations as those presented earlier in Gross and Harris. It follows that $W_q(t)$ may be written as

$$W_q(t) = 1 + k_1 e^{-.8779t} + k_2 e^{-(3.811+.5335i)t} + k_3 e^{-(3.811-.5335i)t}.$$

In order to guarantee real values for $W_q(t)$, k_3 must be the complex conjugate of k_2 . The result can then be simplified to

$$W_q(t) = 1 + k_1 e^{-.8779t} + k_2 e^{-(3.811+.5335i)t} + \bar{k}_2 e^{-(3.811-.5335i)t}.$$

When the work of Smith is applied and the corresponding residues obtained, we find that $k_1 = -.5141$ and $k_2 = .02884 + .03575i$. The final expression for the stationary waiting-time CDF is thus

$$W_q(t) = 1 - .5141e^{-.8779t} + e^{-3.811t} [.05768 \cos .5335t - .07150 \sin .5335t].$$

V. CONCLUDING REMARKS

It should be noted that several other methods are available in the literature which provide similar approximating forms for the service-time distribution. In particular, Luchak (1956) and Wishart (1959) expanded the Stieltjes transform $B^*(s)$ in a Laurent series about $(-\mu)$, and then truncated after k terms to give

$$B^*(s) = \sum_{m=1}^k c_m (1+s/\mu)^{-m}.$$

(This corresponds to approximating $B(t)$ as a mixture of Erlangs with the same scale.) But these $\{c_m\}$ are not likely to sum to one, and thus the truncation is not a true CDF. We may try instead to match moments and use boundary conditions for estimating μ and the $\{c_m\}$. Unfortunately, the latter estimates may not be the same as those from the Laurent expansion and would not necessarily possess any of the major properties typically desired. Observe that such Laurent forms are indeed members of the class R_n .

In closing now, we believe it important to view this work as a total package: numerical, statistical and mathematical. We see our approach as a unification of some of the most fundamental results on the GI/G/1 queue, particularly building on the early and basic work of Smith. This and related work which have followed are all put in total perspective by Cohen. The numerical portions of this study began with the optimization procedure and ultimately led to the derivation of the stationary delay CDF. But we really cannot separate this from the attempt to find quality statistical estimators. By virtue of the fact that the distribution selection is performed by MLE, it is clear that this sort of complete modeling effort is very much data oriented and therefore a potentially powerful tool for the applied queueing analyst.

Finally, the extreme simplicity of the mathematical solution to the waiting-time problem should be reemphasized.

REFERENCES

- J. W. Cohen, The Single Server Queue (2nd ed.), Amsterdam: North Holland, 1982.
- M. Dehon and G. Latouche, "A Geometric Interpretation of the Relations Between the Exponential and Generalized Erlang Distributions," Adv. Appl. Prob., Vol. 14 (1982), pp. 885-897.
- A. A. Fredericks, "A Class of Approximations for the Waiting Time Distribution in a GI/G/1 Queueing System," B.S.T.J., Vol. 61 (1982), pp. 295-325.
- D. Gross and C. M. Harris, Fundamentals of Queueing Theory, New York: John Wiley, 1974.
- A. R. Kaylan, "Statistical Analysis of Finite Mixtures of Exponential and Weibull Mixtures," Ph.D. Dissertation, Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, New York, 1978.
- A. R. Kaylan and C. M. Harris, "Efficient Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures," Computers and Operations Research, Vol. 8 (1981), pp. 97-104.
- G. Luchak, "The Solution of the Single-Channel Queueing Equation Characterized by a Time-Dependent Poisson-Distributed Arrival Rate and a General Class of Holding Times," O.R., Vol. 4 (1956), pp. 711-732.
- J. Mandelbaum, "Numerical Issues in Inference for Finite Mixtures of Weibull Distributions Under Progressive Censoring," Doctoral Dissertation, Department of Operations Research, The George Washington University, Washington, D.C., 1982.
- J. Mandelbaum and C. M. Harris, "Parameter Estimation Under Progressive Censoring Conditions for a Finite Mixture of Weibull Distributions," Studies in the Management Sciences, Vol. 19 (1982), pp. 239-260.
- W. G. Marchal and C. M. Harris, "A Modified Erlang Approach to Approximating GI/G/1 Queues," J. Appl. Prob., Vol. 13 (1976), pp. 118-126.
- A. W. Naylor and G. R. Sell, Linear Operator Theory in Engineering and Science, New York: Holt, Rinehart and Winston, 1971.
- M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models, Baltimore: Johns Hopkins, 1981.
- W. L. Smith, "On the Distribution of Queueing Times," Proc. Camb. Phil. Soc., Vol. 49 (1953), pp. 449-461.
- D. M. G. Wishart, "A Queueing System with Service-Time Distribution of Mixed Chi-Square Type," O.R., Vol. 7 (1959), pp. 174-179.

DISTRIBUTION LIST

Copy No.

1	Office of Naval Research 800 North Quincy Street Arlington, VA 22217 Attention: Group Leader, Statistics and Probability Mathematical and Physical Sciences
2	ONR Resident Representative Joseph Henry Building, Room 623 2100 Pennsylvania Avenue, N.W. Washington, DC 20037
3 - 8	Director, Naval Research Laboratory Washington, DC 20375 Attention: Code 2627
9 - 20	Defense Technical Information Center Building 5, Cameron Station Alexandria, VA 22314
21 - 22	C. M. Harris
23 - 24	E. H. Pancake Sci./Tech. Information Center
25	RLES Files

4470:jt

UNIVERSITY OF VIRGINIA
School of Engineering and Applied Science

The University of Virginia's School of Engineering and Applied Science has an undergraduate enrollment of approximately 1,500 students with a graduate enrollment of approximately 500. There are 125 faculty members, a majority of whom conduct research in addition to teaching.

Research is a vital part of the educational program and interests parallel academic specialties. These range from the classical engineering disciplines of Chemical, Civil, Electrical, and Mechanical and Aerospace to newer, more specialized fields of Biomedical Engineering, Systems Engineering, Materials Science, Nuclear Engineering and Engineering Physics, Applied Mathematics and Computer Science. Within these disciplines there are well equipped laboratories for conducting highly specialized research. All departments offer the doctorate; Biomedical and Materials Science grant only graduate degrees. In addition, courses in the humanities are offered within the School.

The University of Virginia (which includes approximately 1,500 full-time faculty and a total full-time student enrollment of about 16,000), also offers professional degrees under the schools of Architecture, Law, Medicine, Nursing, Commerce, Business Administration, and Education. In addition, the College of Arts and Sciences houses departments of Mathematics, Physics, Chemistry and others relevant to the engineering research program. The School of Engineering and Applied Science is an integral part of this University community which provides opportunities for interdisciplinary work in pursuit of the basic goals of education, research, and public service.

END

FILMED

4-84

DTIC